DOCUMENT RESUME

ED 346 773 HE 025 579

AUTHOR Narveson, Robert D.
TITLE Diversity in American Academic Culture: Implications for Reward Structures.
PUB DATE 23 Apr 92
NOTE 26p.; Paper presented at the American Educational Research Association Symposium on Rewarding Teaching at Research Universities (San Francisco, CA, April 23, 1992).
PUB TYPE Reports - General (140) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Faculty; Documentation; *Evaluation Methods; Higher Education; Incentives; Personnel Evaluation; Recognition (Achievement); Rewards; Student Evaluation of Teacher Performance; *Teacher Evaluation; *Vocational Evaluation
IDENTIFIERS American Educational Research Association; *University of Nebraska Lincoln

ABSTRACT

This paper addresses the problem of how teaching performance should be evaluated and rewarded and examines how a University of Nebraska (Lincoln) Project FIPSE (Fund for the Improvement of Postsecondary Education) has helped to improve the climate for teaching and the processes for recognizing and rewarding excellence. First, the paper examines the sources of diversity that were found at the five colleges involved in the project. The differences described include areas such as department sizes, differences in their missions, and differences in administrative structures and department norms. Discussed next are how, in the face of all the diversity found, appraisal uniformity can be accomplished. Several appraisal practices are compared, and appraisal instruments, the use of student evaluation forms as well as departmental and school evaluation forms, and the appraisal hierarchy that is in place are discussed. Specific conclusions are provided regarding the rationales for appraisal practices, the effects of appraisal practices, and the results of the FIPSE project. A more general conclusion is that the project has demonstrated that when diversity of standards, norms, and values within units is acknowledged and respected, and when units, with administrative encouragement, devise evaluation processes acceptable to their faculties, improvement of both the climate for teaching and the process of evaluation and reward can be achieved. Contains eight references. (GLR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED346773

Robert D. Narveson
Department of English
University of Nebraska
Lincoln, NE 68588-0333

Diversity in American Academic Culture: Implications for Reward Structures

I. The problem in rewarding teaching: Inadequate (Ineffective) Incentives.
- A. A Universal Problem
- B. Diverse Causes
- C. To Make Headway

II. Sources of Diversity.
- A. Size of unit
- B. Mission
- C. Administrative Structure
- D. Norms
  1. Autonomy
  2. Availability
  3. Accountability
  4. Advice
  5. Assignment
  6. Personal Values

III. Appraisal Practices
- A. Relation to Faculty Efforts.
- B. The Appraisal Hierarchy.
- C. Appraisal Instruments.
- D. IANR Student Evaluation forms.
  1. College-wide forms vs. Departmental forms.
  2. The CASNR form examined.
  3. Rationale for Uniformity.
- E. Arts and Sciences forms examined.
  1. Political Science.
     a. Reflection first.
     b. Many questions.
  2. English: few questions.
  3. Music: Multiple forms.
  4. The Desire for Autonomy.

IV. Conclusions.
- A. Rationales for Appraisal Practices.
  1. Reward for Accomplishment.
  2. Incentive, through merit awards, to improve teaching.
- B. Effects of Appraisal Practices
  1. Probably Dysfunctional
  2. Evaluation alone: little improvement
  3. Consultation and Improvement.
- C. Results of FIPSE Project
  1. Evaluation for Improvement
  2. Collegial Cooperation
  3. Teaching Portfolios.
  4. Self-appraisal and Re-evaluation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert D. Narveson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Robert D. Narveson
Department of English
University of Nebraska
Lincoln, NE 68588-0333

Diversity in American Academic Culture: Implications for Reward Structures.

Paper for Symposium on Rewarding Teaching at Research Universities, AERA, San Francisco, April 23, 1992.

I. The Problem: Inadequate (Ineffective) Incentives.

A. A Universal Problem. Responding to widespread complaints that undergraduate teaching is neglected, particularly at research universities, critics have suggested that reward structures offer inadequate incentives to teachers to devote time to teaching activity and teaching improvement. There is little agreement, however, on how teaching performance should be evaluated and rewarded in order to provide effective incentives.

B. Diverse Causes. How to evaluate and reward teaching performance is not, however, a problem lending itself to a simple solution. While much in institutional culture in the United States is similar everywhere, and the similarities must be taken into account, the differences are at least as great. Differences must be taken into account, both the differences in the people who do the teaching, and differences in their institutional settings, for these differences are causes of problems in performance appraisal.

C. Making Headway. By recognizing and respecting diversity, the University of Nebraska FIPSE Project on Rewarding Teaching has made headway toward improving the climate for teaching and the processes for recognizing and rewarding excellence.

II. Sources of Diversity.

Leaving to psychologists the question of what makes individual people different, let us focus on what makes departments, colleges, and institutions different. Dimensions along which variations occur are easily enumerated, though terms of classification may differ. Blackburn and Pitney define important dimensions under the headings of structures, norms, values, and practices. Structures, norms and values, they say, are similar across postsecondary institutions, "but how performance appraisal is practiced certainly is not" (26). If structures, norms, and values are similar, what causes the dissimilarities in appraisal practices? Our work with twenty eight departments in five colleges at the University of Nebraska at Lincoln has shown us that subtle and not so subtle differences of many kinds exert strong influences on practices.

A. Size, obviously. We have worked with departments with as little as 1.5 FTE (Plant Pathology) and as many as 50 FTE (English). Smaller departments, such as Industrial Engineering, tell us that they operate most comfortably with informal, unwritten procedures. Large departments such as English or Agronomy, tend to have their procedures spelled out in detail in written documents. Size alone forces differences in practices.

B. Mission. We are working with one department with only graduate students (Adult Restorative Dentistry), one with 90 percent graduate students (Plant Pathology), departments with 90+ percent undergraduate students (Mathematics), and departments with no graduate students and no classes above the second-year level (Agricultural Communications). We work with departments in which service courses for first year and second year students comprise 80

percent of their credit hour production (English, Math, Psychology, History). The relationship of teaching to scholarly interests may be very different for a person teaching a generalized introductory course and for someone teaching a course for majors. On the other hand, in some disciplines, such as English composition or beginning foreign language courses, the chief interest of the discipline, and therefore the chief opportunities for scholarly publication, may be in the area of teaching. All of these differences make a difference in how units approach the problems of evaluation and reward.

C. Administrative Structure. Colleges are also very different in their relationship to departments. In the Institute of Agriculture and Natural Resources--of which the College of Agricultural Sciences and Natural Resources is the teaching unit--administration tends to be more centralized than in the College of Arts and Sciences. The consequence is to give an impression that the former college knows more clearly what it is about than does the latter--not that is in all cases a good thing. This centralization of power and control in IANR is reflected in separation of teaching, research, and service missions into different offices with separate budgets under different deans. There is a Dean of Instruction (Don Edwards), a Dean of Research (Darryl Nelson), a Dean of Cooperative Extension (Leo Lucas), and a Dean of International Programs (Jean Eigner), each with a budget that pays a set percentage of a faculty member's salary, reflecting that person's assignment. Centralization of power is reflected in titles of leadership in departments ("heads" as opposed to "chairs") and in the practice of mandating a single student and course evaluation form to be used throughout the Institute. In the College of Arts and Sciences, by contrast, a single Dean (John Peters)

administers a single budget and is responsible for overseeing teaching, research, and service performance.

D. Norms.

1. Autonomy. The degree of autonomy granted by the institution to each college results in differences in administrative practices. It is a peculiarity of the University of Nebraska that under its governing laws, each college a degree of autonomy in setting its requirements for graduation, and in fact in Engineering and Technology, the courses that meet distribution requirements are set by the departments individually.

2. Availability. The prominence in the Institute of its service mission to agriculture is reflected in the availability of the faculty, whose whereabouts is nearly always known, in contrast to the unavailability of the faculty in Arts and Sciences, whose whereabouts is known only during scheduled class and office hours. Reaching a member of the Institute faculty by phone is simple; reaching a member of the Arts and Sciences faculty by phone is chancy.

3. Accountability. Another difference in norms is evident in traditions of accountability. In some departments, classroom visits by the chair or other supervisor are standard and accepted. In other departments, the chair never visits classes, and faculty do not feel comfortable about visiting each other's classes. In some departments there is a tradition of senior professors serving as mentors, formally or informally. The mentoring is most commonly done to help the junior faculty member with his/her research

program. It is less commonly teaching-related. Informal exchange of visits to classes occurs in some departments; in others it is rare or unknown. It is safe to say that at the University of Nebraska there is no normative practice.

4. Advice. Many institutions now have an office of faculty development (at Nebraska it is called the Teaching and Learning Center). These centers provide help to classroom teachers who feel incentive to improve. In some departments there is a high degree of confidence in the help offered by the TLC; in others, the TLC is little valued. Where chairs refer to colleagues perceived to have problems with teaching to the TLC, one effect is to remove responsibility for improvement of classroom performance from colleagues in the department. Some departments reflect their sense of collegial responsibility for improvement of teaching through a formal arrangement by which colleagues assist each other in examining and improving their teaching practices. Psychology calls its arrangement "Teaching Circles" and Agronomy has a system of five year consultation and review by a committee of peers.

5. Assignment. In the Institute for Agriculture and Natural Resources individual assignments differ, reflecting the mission, norms and values of the Institute. In Agronomy, one professor (Bob Sorensen) has an assignment of 100 percent teaching, and other professors have 100 percent research and service assignments. In the College of Arts and Sciences, on the other hand, every professor is assumed to have responsibilities in all three areas of teaching, research, and service, though the percentage of effort

assigned for each area may vary widely, from as little as 20 percent teaching to as much as 60 percent teaching.

6. Personal values. At an institution where faculty are hired with the expectation that they will produce publishable research, and where promotion to full professor is normally achieved only through a substantial record of published research, a high proportion of professors think of themselves first and foremost as researchers. Nevertheless, another large proportion think of themselves first and foremost as teachers, and dedicate their energies not merely to their own teaching but also to the teaching mission of the unit. The greatest successes of the Nebraska project were achieved in those units in which the chairs identified and enlisted faculty members to serve as departmental coordinators who were not only respected publishing scholars but were also respected for their dedication to and success at teaching.

III. Appraisal Practices. Differences along so many dimensions as these leads to great diversity in appraisal practices. Such diversity makes any uniformity of practices in evaluation and reward both unlikely and potentially disfunctional. But while uniformity is not necessarily a desirable goal, appraisal itself is unavoidable. How then is it best done?

A. Relation to Faculty Efforts. One assumption of the University of Nebraska FIPSE Project on Rewarding Teaching has been that a well-designed reward structure significantly affects the activities of the faculty. If teaching and teaching-related activity are overtly rewarded and are seen to be

rewarded, those already by inclination dedicated to teaching have an additional motive for dedicating their energies in those directions, while those less inclined may be led to reconsider their distribution of energies. Problems with this assumption are discussed later.

B. The Appraisal Hierarchy. Structures, norms, values, plus external pressures ("determinants"--Blackburn and Pitney 26), influence the appraisal practices of each unit. Appraisal practices for reward of teaching include 1) documentation, usually collected by the teacher, and 2) assessment, usually conducted at multiple levels: first by students through course- and instructor-evaluations, then by peers through personnel committees, then by chairs or heads, then by college deans and executive committees, then, at least in theory, by vice chancellors, chancellors, president, and the Board of Regents. At every point, differences in structures, norms, and values affect the practices of assessment and reward in individual cases. While departmental norms and values are the most powerful in departmental decisions, departments must temper their own judgments by considerations of norms and values of higher units that will review their decisions. Whatever the theoretical hierarchy involved, our deans tell us that assessment in the departments is most critical, yet it is among departments that diversity of structures, norms, and values is greatest.

C. Appraisal Instruments. One article cannot investigate all the sources of variation in appraisal practices. I will narrow the topic by discussing only appraisal instruments used in evaluation of teaching, and within those, will take a detailed look, for illustrative purposes, only at

student evaluation forms. In this paper I define "appraisal instruments" as the written documents used by the colleges and by the departments used in appraisal of professional performance. In appraising faculty performance, the departmental chairs or heads, and the committees of peers, where they are involved, use an annual report prepared and submitted by each faculty member. The report typically includes the filled-out student evaluation forms from courses of the past year, a summary and averaging of scores on those forms, plus, increasingly, a variety of other information concerning both classroom and non-classroom teaching-related activities. The chair or head summarizes in a letter to the dean the assessments by peers, the evidence from the student evaluations, plus any other evidence, and recommends merit and other salary increases for the coming year.

D. Student Evaluation Forms. The subtle differences among units in structures, norms, and values may be inferred by examining student evaluation forms used in units of the University of Nebraska at Lincoln that participated in our study. We can reasonably attempt to infer from these appraisal instruments some "messages" about norms and values that these instruments deliver. Our conclusions will necessary be speculative, but no less interesting because of that. Our primary conclusion will be that the sum total of all these cumulative differences in structures, norms, values, and practices guarantees that no one scheme of evaluation and reward will be acceptable to the variety of units in an American research university such as ours.

1. College-wide forms vs. Departmental forms. The questionnaires used in gathering student evaluative data differ from department to department and from college to college. The College of Agricultural Science and Natural Resources has long had a standard form used by all departments. The latest version has twenty one machine-scored questions and seven questions inviting written comments. Use of this form is required in all departments each semester. In the College of Arts and Sciences there is no official form, the only college requirement being that "a departmental or school file is kept for each faculty member to record . . . evidence of teaching effectiveness (for example, by means of student evaluations)" (A&S Faculty Handbook 20). In practice this has come increasingly to mean that student evaluations are administered in each course at least once each year. In some departments, such as Mathematics, a common form is mandated for all instructors. In others, such as History, though a departmental form is available, instructors frequently choose or devise their own forms.

2. The CASNR form. The College of Agricultural Science and Natural Resources (CASNR), in keeping with its structure, norms, and values, requires the same form of everyone. In so far as the Ag College form expresses an ideal of good teaching, the emphasis is on the organization and content of knowledge. There are four questions that imply a desired content and organization of knowledge.

9. The instructor did NOT synthesize, integrate, or summarize effectively.
11. The instructor encouraged development of new viewpoints and appreciations.
13. Some things were NOT explained very well.

14. The instructor demonstrated a thorough knowledge of the subject matter.

The implications of these items are that knowledge is hierarchical (item 9), incomplete (item 11), available (13), and there in an objective sense (13 and 14). As Tom C. Wilson points out, forms carry with them the authority of their authors. "'[P]roper' pedagogy is subtly dictated by the unnamed author or authors of the form" (83). He argues that evaluation forms typically imply that "the truth of the world is discovered through empirical observation, and disagreement about what constitutes good practice can be decided by collecting and analyzing facts. Good teaching can be reduced to behaviors based on a tight experiment" (84). These assumptions about knowledge he characterizes as positivistic. Such assumptions imply roles for teacher and student in which the teacher passes discrete knowledge to the students by arranging and packaging it in assimilable bits.

My point is not that these assumptions are faulty; rather it is that other departments operate from other assumptions, and use evaluation forms reflecting the differences.

3. Rationale for uniformity. Tom C. Wilson (1988) assumes that the use of a uniform evaluation form for the whole department, college or institution reflects an administrative desire for comparative data. Common sense and experience justify such an assumption. A unit-wide form usually includes items that can be scored numerically and averaged. "If only write-ins were allowed, responses might not be uniform enough for statistical interpretation" (Wilson 82). Comparisons across units is a use which faculty

members and departments within the College of Agriculture must and do accept.

E. Arts and Sciences Evaluation Forms. By contrast, faculty members and departments in the College of Arts and Sciences reject a common form along with the virtue of cross-unit comparisons. The forms used in departments vary from those with a family resemblance to the lengthy IANR form to those with as few as five generic questions.

1. The Political Science Department of the College of Arts and Sciences has a form resembling the Institute form in its use of a considerable number (13) of questions calling for scaled ratings of instructor and course. Interestingly, though, the Political Scientists differ from most departments and colleges in placing their requests for written comments about instructor and course first, and the questions asking for numerical ratings second. In five separate items, students are requested to think about and describe specific examples of things liked and disliked about instructor and course, specific suggestions for improvement of instructor and course, and specific ways in which the course or instructor differed from the student's expectations.

What is the thinking behind this reversal of the more common practice? One can only speculate that the department hopes by this practice to receive more reflective and considered responses to the questions that ask for scaled ratings. Possibly this reflects a belief that an intellectually responsible person bases ratings of a teacher on conscious reflection more than on "gut" reactions. Support for this explanation may be sought in the wording of

Question 11, which asks the student to rate "the degree to which the course helped you to develop analytical skills, such as thinking, analyzing and expressing yourself clearly."

In another difference from the IANR form, the Political Science form recognizes that student response to a course may be heavily determined by factors unrelated to the course. Question 12, which asks the student to rate "the amount you learned (whether knowledge or skills) in this course compared with other courses at the same level you have had at UNL," concludes with this parenthetical sentence: If you feel you didn't learn much due to YOUR lack of effort or personal problems, mark N/A. Perhaps more than any other departmental form, the Political Science form in this way attempts to recognize and impress on the student evaluators that learning is at least a two-way street.

2. English. Faculty members knowledgeable in statistics tend to distrust the statistical interpretation of questionnaire data and may offer persuasive arguments for their distrust (Amedeo). Disciplines without sophisticated knowledge of statistics, such as English, sometimes place faith in statistical averages, even while stating an explicit distrust. This faith is then tempered by a subjective reading of student comments. The distrust of statistical data from small classes may be expressed in other ways as well, such as in a questionnaire with relatively few questions, and those of a global nature.

The form used by the English Department in the College of Arts and Sciences consists of five questions to which students are asked to rate the strength of their agreement or disagreement on a continuum from one to five:

1. This is an excellent course.
2. This is an excellent teacher.
3. The instructor encouraged me to do my best work.
4. I am performing up to my potential in this course.
5. I had a strong desire to take this course.

There are in addition four questions asking for student comments on course, self, instructor, and "anything else."

The English department has at various times experimented with lengthier forms that inquire into particular aspects of instruction, but has generally concluded that the information gathered through additional, more particular questions is not useful to the personnel committee, and may be taken to imply a party line concerning what are desirable teaching practices. The department does not wish to suggest that one mode of instruction or one set of instructional practices is best for all teachers or all classes, and trusts the individual teacher to exercise professional judgment about how best to teach. Low scores on student evaluations are taken as evidence of something amiss.

Why are the questions so few? The department accepts, officially and in practice, the idea that, on the whole, student judgments of teaching

excellence are accurate. Members of the personnel committee boast and complain about the time they spend reading every student comment. Because the numbers are small, and because English teachers are uncomfortable with numbers, the chair and the committee of peers that ranks teachers for merit raises read the student comments and adjust the numbers according to their interpretation of what the comments indicate. They arrive at a holistic judgment concerning what the evaluations show.

The contrasting philosophies that underlie the English department form and the IANR form could be summed up like this. The English department distrusts the crude choices forced on people by multiple choice questions, and places more faith in explanatory comments generated by people in explaining their value judgments. The emphasis is on the uniqueness and autonomy of individuals in ordering and valuing their experiences. The IANR form reflects an emphasis on generalizations abstracted from individual ordering and valuing. The emphasis is on community consensus about what is knowledge and what is valued by the community. Individual behavior is subordinated to group expectations.

An open-ended form such as that used in English leaves the instructor in the dark about what the readers of the student comments will look for and value. The person evaluated has only a vague sense of what specific matters influenced the final assessment. A more explicit form such as used in the Institute allows instructors being evaluated a certain degree of knowledge about what the readers of the student comments will be looking at and for.

3. The School of Music. The English department and most departments in the College of Agriculture do their teaching primarily in classes. The School of Music, the College of Dentistry, and the Art Department, do much of their teaching in laboratory or studio settings, and in one-on-one coaching sessions. These units have developed student evaluation forms appropriate to each form of instruction practiced in their units. In Music the forms differ for Applied Music (individual instruction), Ensemble, Opera/Music Theater, Academic Courses, and Methods and Skills Courses.

Students in Applied Music are asked to check blanks in columns labeled "Outstanding, Excellent, Average, Acceptable, Poor, or Unable to Judge" or "Always, Often, Seldom, Never, Unable to Judge." Unlike most other forms, the Music forms do not assign numbers to these categories, but the ease of assigning numbers to each category is evident. Whether or not the inference is accurate, a distrust of numbers may be inferred.

Judgments are asked for on topics in three separated categories: Repertoire, Instructor's Knowledge and Manner, and Instructor's Interest and Concern. Students are asked for frequency of specific types of actions by the instructor. For example, Was the instructor attentive during lessons? Did the instructor make specific recommendations for improvements? The forms for other instructional settings and projects in the School of Music are parallel in structure but change some of the questions to reflect differences in instructional mode. The overall thrust in the School of Music is to recognize and specify the differing demands made on teachers by different instructional

settings, and fear of being judged by inappropriate standards by uncomprehending judges is easily inferred.

4. Desire for Autonomy. Perhaps more revealing for Arts and Sciences College student evaluation forms than the content of the forms is that each department insists on its own form adapted to its own purposes. Considerations such as those described above are weighed very differently by different units, and units resist vigorously any attempt to impose forms or practices of which they do not approve. The purposes may be justified by different rationales; one must wonder, however, whether the real significance is what it says about each department's desire to think of itself as autonomous. This autonomy is far from complete in reality, for the money to run the department and pay salaries comes to each department, at least formally, at the discretion of the dean of the college. Departments must satisfy the dean that they are adequately monitoring the performance of faculty members. Departments have free will so long as they don't try to exercise it in ways offensive to the dean. Throughout the College of Arts and Sciences, however, it is clear that there is a strong desire in each unit to think of itself as unique and therefore deserving and requiring autonomy. A telling indicator of assumption of autonomy may be the variety of dress and absence of a dress code among faculties of the College of Arts and Sciences. By contrast, a fairly strong uniformity of dress is practiced by faculties of IANR.

IV. Conclusions. The differences in student evaluation forms reflect the wide variety of norms, values and structures among units. If we looked at other

3. The School of Music. The English department and most departments in the College of Agriculture do their teaching primarily in classes. The School of Music, the College of Dentistry, and the Art Department, do much of their teaching in laboratory or studio settings, and in one-on-one coaching sessions. These units have developed student evaluation forms appropriate to each form of instruction practiced in their units. In Music the forms differ for Applied Music (individual instruction), Ensemble, Opera/Music Theater, Academic Courses, and Methods and Skills Courses.

Students in Applied Music are asked to check blanks in columns labeled "Outstanding, Excellent, Average, Acceptable, Poor, or Unable to Judge" or "Always, Often, Seldom, Never, Unable to Judge." Unlike most other forms, the Music forms do not assign numbers to these categories, but the ease of assigning numbers to each category is evident. Whether or not the inference is accurate, a distrust of numbers may be inferred.

Judgments are asked for on topics in three separated categories: Repertoire, Instructor's Knowledge and Manner, and Instructor's Interest and Concern. Students are asked for frequency of specific types of actions by the instructor. For example, Was the instructor attentive during lessons? Did the instructor make specific recommendations for improvements? The forms for other instructional settings and projects in the School of Music are parallel in structure but change some of the questions to reflect differences in instructional mode. The overall thrust in the School of Music is to recognize and specify the differing demands made on teachers by different instructional

ascertain from students their level of satisfaction with the instruction they have received.

2. Incentive, through Merit Awards. The principle underlying merit systems is that superior work should receive superior reward. In addition, to rewarding past performance, however, taking good teaching into account in setting salaries supposedly provides an incentive to teachers to improve their performance. But do rewards for good teaching lead to improvement in teaching? Unfortunately, there is little or no evidence to support the claim that they do, and some evidence to contradict it.

B. Effects of Appraisal Practices.

1. Probably Disfunctional. What is problematic about appraisal, according to Blackburn and Pitney, is that "[performance appraisal of faculty] is probably dysfunctional. The haves and the have-nots grow farther apart each year, and few performance improvements are realized. Social support goes down, along with morale. Alienation exacerbates the problem. When this happens, the institution and the individual faculty are less likely to accomplish their goals" (29).

2. Evaluation Alone: Little Improvement. If the goal of evaluation is to improve teaching, disappointment and disillusion is a potential result. "[S]tudies show[ed] that in the absence of intervention, faculty members generally show little change in the ratings they receive from students about their teaching effectiveness from one time to another" (Wilson [1986] 208). Wilson cites studies by Rotem and Glasman, and Kulik.

Blackburn and Pitney (27) also conclude: "In our case studies we learned that student evaluations had inconsequential effects on faculty teaching behavior. . . . First of all, most receive high scores. 'Why change?' they ask. Second, and more important, almost without exception faculty give less than full credence to student evaluations, and some consider them worthless." No wonder faculty members tend to distrust incentives based on student evaluations!

3. Consultation and Improvement. When evaluation for formative purposes is properly used, however, studies have shown that it can be effective. In his report of his project at Berkeley, Wilson concluded that when teachers worked with trained consultants, they were able to use their student evaluations to improve their teaching. "We found that when departments surveyed their own attitudes and reviewed the research findings that we provided for them, many of them did conclude that efforts by instructors to improve their teaching, through consultation and by other means, were of value." And if that is the case, efforts to improve, they concluded, are worthy of reward.

C. Results of FIPSE Project. With considerations such as these in mind, units at the University of Nebraska in Lincoln have been able to design improvements in their appraisal practices.

1. Evaluation for Improvement: UNL. In the light of research findings such as those reported by Robert C. Wilson, and other studies cited by Blackburn and Pitney, an important variable in institutional practice may be the uses made of student evaluation data to improve teaching performance.

What use is being made of such data in units at the University of Nebraska? Peter Seldin and others have emphasized that information may be gathered from students for two very different purposes: to inform personnel decisions (summative judgment), and to assist instructors in improving their teaching (formative judgment). Seldin further argues that information needed for summative purposes is different from information needed for formative purposes. The questionnaires examined all request from students information that goes beyond asking how well students think the instructor has performed. In some cases, information useful for formative purposes is used for summative purposes, increasing faculty distrust of student evaluations. Though the departmental chair or head may advise individuals to seek help from colleagues or from the Teaching and Learning Center, there is not yet general acceptance as of yet of a collegial responsibility to help teachers improve. As a result, there is as yet no general practice of rewarding teachers for efforts to improve their teaching, and on the whole the reward system has not functioned well as a mechanism for the improvement of teaching. This is an area in which units are experimenting, with positive results.

2. Collegial Cooperation. Greater collegial cooperation for improving teaching is an area in which the Project on Rewarding Teaching is making a documentable difference. A few participating departments have begun to establish formal means to recognize and reward teachers for working with colleagues to institute practices that result in student satisfaction and to change classroom practices that result in student dissatisfaction. The Psychology Department's provision for teaching circles is being imitated in a number of departments, as is the Agronomy Department's establishment of a five

year review for teacher and course by a committee of peers. The English department has initiated steps to put together an idea book of the sort reported in the Berkeley study and to encourage its use by instructors in consultation with colleagues who visit each others' classes, examine each others' teaching materials, and read each others' student evaluations. Progress is being made and there is still much to do.

3. Teaching Portfolios. As one would expect from what has been said so far, departments in the Nebraska FIPSE project differed in what they thought were rewardable improvement activities. But most moved in the direction of a teaching portfolio as recommended by Seldin, by Blackburn and Pitney, and by many others. Student evaluations form one portion, but only a portion, of a teaching portfolio. Many of our departments see the portfolio concept as offering a flexible approach to documenting performance. Advantages of a portfolio approach to evaluation are summarized by Blackburn and Pitney: "We believe a portfolio system would accomplish the goal of continuous faculty growth and development, the realization of the individual's full potential. We believe that the portfolio process fits the conclusions from the research literature and college and university norms, values, structures, and practices. . . . By definition, the portfolio process is individualized. . . . The portfolio process can directly address the discipline and specialty differences on kinds of products and rates of production. . . . In addition, the process involves the individual being evaluated and certainly could involve peers. . . . The portfolio process also allows for the separation of attention to growth and development from merit and salary increases, or can for those faculty who would prefer the two

aspects to be considered separately. . . . Portfolios ought to improve the organization as well as benefit the individuals" (32-33).

Units that adopt a teaching portfolio approach to documenting teaching performance still face the task of learning how to assess portfolios to determine degree of merit.

4. Self-appraisal and Re-evaluation. Even in departments that show little documentable change in their practices, their testimony is that the scrutiny given to teaching reward resulted in public discussion and private rethinking and reworking of teaching practices. The ground even in these departments has been prepared for more substantive changes later on. For example, the forms in some departments solicit data about the student doing the evaluating: year in school, number of terms of study with this instructor, degree area, number of sessions missed, grade student expects, and grade student thinks is deserved. The clear implication to the student is that who the student is and how the student has behaved will influence the evaluations the student gives and that those who read the evaluations will interpret both the evaluations and the evaluators. Mathematics has in the past requested such information. The chair of Mathematics recently informed me, though, that it will no longer do so, because, having considered the matter, it sees no way to use the information in assessment as practiced by the department. This conclusion reached by the mathematics department illustrates an important point: The chief benefits derived from a project such as ours will not be the production of written documents detailing procedures for evaluation and reward. Self-appraisal by departments may well be the most important single result of the project.

We have now worked with twenty eight different departments in five different colleges. While their blueprints for evaluations and rewards vary widely, and from different points of view may fairly be judged better or worse, each participating department now knows what other departments are doing and has to justify and defend its own plans. We have seen that raised consciousness has produced better evaluations of teaching and better rewards for teaching in a number of cases. That this has happened reflects the activity not only of the departments but of the deans of the colleges. To some extent, perhaps, a greater uniformity of practices is evolving. More importantly, however, the project has demonstrated that when diversity of standards, norms and values within units is acknowledged and respected, and when units, with administrative encouragement, devise evaluation processes acceptable to their faculties, improvement of both the climate for teaching and the process of evaluation and reward can be achieved.

Works Cited

Amedeo, Doug. "Some Comments on the Data Utilized for the Department's Merit Matrix and Its Evaluation." Department of Geography, University of Nebraska, Lincoln, NE 68588. Unpublished (1991). Cited with permission.

Blackburn, Robert T. and Judith A. Pitney. Performance Appraisal for Faculty: Implications for Higher Education. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, 1988.

Kulik, J. A. "Student Reactions to Instructions," in Teaching in Higher Education, ed. by S. C. Scholl and S. C. Inglis. Ohio Board of Regents, 1977.

Ory, John C. "A Role for Assessment in Higher Education Decision Making." In P. J. Gray (ed). Achieving Assessment Goals Using Evaluation Techniques. New Directions for Higher Education, No. 67. San Francisco: Jossey-Bass, Fall 1989, pp. 71-87.

Rotem and Glasman. "On the Effectiveness of Students' Evaluative Feedback to University Instructors," Review of Educational Research, 49 (1979), 497-511.

Seldin, Peter. Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass, 1985.

Wilson, Robert C. "Improving Faculty Teaching: Effective Use of Student Evaluations and Consultants." Journal of Higher Education. 57, No. 2 (March/April 1986.

Wilson, Tom C., "Student Evaluation of Teaching Forms: A Critical Perspective." Review of Higher Education, 12: 1 (Fall 1988) 79-95.